Tutoriat PS 3: Dependence and Independence in Discrete Random Variables

Mihai Duzi Lucan Cristian Luminaru Ionuț

Contents

1 Dependence and independence for random variables

1.1 Definition and Terminology

Until now, we made the assumption that our (discrete) random variables are independent. Now we will define what that means.

Two random variables X and Y are independent $\iff$ $p(x, y) = P(X = x, Y = y) = P(X = x) \times P(Y = y) \iff p(x|y) = P(X = x|Y = y) = P(X = x)$

$p(x, y) = P(X = x, Y = y)$ **is called the joint probability mass function.**

Just like before, it has the following properties:

$$0 \leq p(x, y) \leq 1$$

$$\sum_x \sum_y p(x, y) = 1$$

We also have the joint cumulative distribution:

$$F(x, y) = P(X \leq x, Y \leq y)$$

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j)$$

The individual probabilities of $p(x)$ and $p(y)$ are called **marginal probabilities**. They formulas come from total probability law:

$$p(x) = \sum_y p(x, y)$$

$$p(y) = \sum_x p(x, y)$$

And the general formula of the mean involving 2 random variables:

$$E[XY] = \sum_x \sum_y p(x, y)xy$$

$$E[X + Y] = \sum_x \sum_y p(x, y)(x + y) = E[X] + E[Y]$$

1.2 Exercises

Let X1 be the probability distribution of the sums of two dices. Let X2 be the probability distribution of a fair dice.

1. Are X1 and X2 independent?

$$X_1 = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} \\ \frac{1}{36} \end{pmatrix}$$

$$X_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

To check for independence, we must assume a relationship between X_1 and X_2. Let's assume $X_1 = D_1 + D_2$ (where D_1, D_2 are two independent dice) and $X_2 = D_1$ (the first die). We check if $P(X_1 = x, X_2 = y) = P(X_1 = x)P(X_2 = y)$ for all x, y. Let's check for $x = 9$ and $y = 3$.

$P(X_1 = 9, X_2 = 3) = P(D_1 + D_2 = 9, D_1 = 3) = P(D_2 = 6, D_1 = 3)$ Since D_1 and D_2 are independent: $P(D_1 = 3, D_2 = 6) = P(D_1 = 3) \times P(D_2 = 6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

Now we check the product of the marginal probabilities: $P(X_1 = 9) = \frac{4}{36} = \frac{1}{9}$ $P(X_2 = 3) = \frac{1}{6}$

$P(X_1 = 9) \times P(X_2 = 3) = \frac{1}{9} \times \frac{1}{6} = \frac{1}{54}$

Since $\frac{1}{36} \neq \frac{1}{54}$, we have $P(X_1 = 9, X_2 = 3) \neq P(X_1 = 9) \times P(X_2 = 3)$.

This counter example alone tells us that these variables are not independent.

2. What is probability of both getting a sum smaller than 8 and getting a number smaller than 3 on the first dice roll?

Solution: We want to find $P(X_1 < 8, X_2 < 3)$. Assuming $X_1 = D_1 + D_2$ and $X_2 = D_1$:

$$P(D_1 + D_2 < 8, D_1 < 3) = P(D_1 + D_2 < 8 \text{ and } D_1 \in \{1, 2\})$$

We can sum the probabilities of the disjoint events:

$$P(D_1 = 1, 1 + D_2 < 8) + P(D_1 = 2, 2 + D_2 < 8)$$

- If $D_1 = 1$: We need $D_2 < 7$. This is true for all 6 outcomes of D_2. The probability for this case is $P(D_1 = 1) = 1/6$. Or, by counting: (1,1), (1,2), (1,3), (1,4), (1,5), (1,6). (6 outcomes)
- If $D_1 = 2$: We need $D_2 < 6$. This is true for $D_2 \in \{1, 2, 3, 4, 5\}$. The probability for this case is $P(D_1 = 2 \text{ and } D_2 \leq 5) = P(D_1 = 2)P(D_2 \leq 5) = (1/6)(5/6) = 5/36$. Or, by counting: (2,1), (2,2), (2,3), (2,4), (2,5). (5 outcomes)

Total favorable outcomes $= 6 + 5 = 11$. Total possible outcomes $= 36$. The probability is $\frac{11}{36}$.

1.3 Conditional mean

$E[X|A] = \sum_x x \cdot P(X = x|A)$

Using the definition of conditional probability, $P(X = x|A) = \frac{P(\{X=x\} \cap A)}{P(A)}$, this becomes:

$$E[X|A] = \frac{\sum_x x \cdot P(\{X = x\} \cap A)}{P(A)}$$

(where the sum is over all possible values x of X).

Let's calculate $E[X_1|X_2 < 4]$ (X1 and X2 from before.

Using the relationship $X_1 = D_1 + D_2$ and $X_2 = D_1$, this is $E[D_1 + D_2|D_1 < 4]$.

The event A is $D_1 < 4$, which means $D_1 \in \{1, 2, 3\}$.

$P(A) = P(D_1 \in \{1, 2, 3\}) = P(D_1 = 1) + P(D_1 = 2) + P(D_1 = 3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$.

By the linearity of conditional expectation:

$$E[D_1 + D_2|D_1 \in \{1, 2, 3\}] = E[D_1|D_1 \in \{1, 2, 3\}] + E[D_2|D_1 \in \{1, 2, 3\}]$$

We solve each part:

First term: $E[D_1|D_1 \in \{1, 2, 3\}]$

Given the condition, D_1 can only be 1, 2, or 3. Since D_1 is a fair die, these three outcomes are equally likely.

$P(D_1 = 1|D_1 \in \{1, 2, 3\}) = \frac{P(D_1=1 \cap D_1 \in \{1,2,3\})}{P(D_1 \in \{1,2,3\})} = \frac{P(D_1=1)}{1/2} = \frac{1/6}{1/2} = \frac{1}{3}$

Similarly, $P(D_1 = 2| \dots) = 1/3$ and $P(D_1 = 3| \dots) = 1/3$.

The expectation is:

$$E[D_1| \dots] = \left(1 \cdot \frac{1}{3}\right) + \left(2 \cdot \frac{1}{3}\right) + \left(3 \cdot \frac{1}{3}\right) = \frac{1 + 2 + 3}{3} = \frac{6}{3} = 2$$

Second term: $E[D_2|D_1 \in \{1, 2, 3\}]$

Since the two dice D_1 and D_2 are independent, the condition on D_1 gives no information about D_2.

Therefore, $E[D_2|D_1 \in \{1, 2, 3\}] = E[D_2]$. The expected value of a single fair die is:

$$E[D_2] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5$$

Result:

$$E[X_1|X_2 < 4] = 2 + 3.5 = 5.5$$

1.4 Correlation and Covariance

Now we will see how dependence is measured between two random variables.

1.4.1 Deducing the correlation coefficient

Let X and Y be two random variables such that $Y = a \times X + b$ (linear dependence)

- $a > 0$ means X and Y are positively correlated
- $a < 0$ means X and Y are negatively correlated
- a near 0 means X and Y are almost independent

$E[X \times Y] = E[aX^2 + bX] = a \times E[X^2] + b \times E[X]$
$E[X] \times E[Y] = E[X] \times (a \times E[X] + b) = a \times E[X]^2 + b \times E[X]$
Subtracting we get $E[X \times Y] - E[X] \times E[Y] = a \times (E[X^2] - E[X]^2) = a \times \text{Var}[X]$
$a = \frac{E[X \times Y] - E[X] \times E[Y]}{\text{Var}[X]}$
We want to represent variance using both X and Y. $\text{Var}[Y] = (a^2) \times \text{Var}[X]$
$\sigma_Y = |a| \times \sigma_X$
$\sigma_X = \frac{\sigma_Y}{|a|}$
$\text{Var}[X] = \sigma_X^2 = \sigma_X \times \frac{\sigma_Y}{|a|}$
Now we have:
$a = \frac{E[X \times Y] - E[X] \times E[Y]}{\sigma_X \times \frac{\sigma_Y}{|a|}}$
$\frac{a}{|a|} = \frac{E[X \times Y] - E[X] \times E[Y]}{\sigma_X \times \sigma_Y}$
$\frac{a}{|a|} = \text{sgn}(a)$ and gives the type of correlation we have (positive/negative)

1.4.2 Covariance of Two Random Variables

Let X, Y be two random variables with $E[X^2]$, $E[Y^2] < +\infty$.

The covariance of the two random variables X and Y is defined as:

$$\begin{aligned}
\text{Cov}(X, Y) &:= E[(X - E[X])(Y - E[Y])] \\
&= E[XY - E[X] \cdot Y - X \cdot E[Y] + E[X] \cdot E[Y]] \\
&= E[XY] - E[X] \cdot E[Y] - E[X] \cdot E[Y] + E[X] \cdot E[Y] \\
&= E[XY] - E[X] \cdot E[Y].
\end{aligned}$$

X and Y are independent $\implies$ Cov(X, Y) = 0
The inverse is not true! Here is an example
Let the random variable X take the values $-1, 0, 1$ with equal probability:

$$X = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Define $Y = X^2$. Then Y takes the values 0 and 1:

$$Y = \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

The joint distribution of X and Y is:

$X \backslash Y$	0	1
-1	0	$\frac{1}{3}$
0	$\frac{1}{3}$	0
1	0	$\frac{1}{3}$

$$E[X] = (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$$

$$E[Y] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$$

$$E[XY] = (-1)(1) \cdot \frac{1}{3} + 0(0) \cdot \frac{1}{3} + (1)(1) \cdot \frac{1}{3} = 0$$

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - (0)\left(\frac{2}{3}\right) = 0$$

Thus, the covariance is zero:

$$\boxed{\mathrm{Cov}(X, Y) = 0}$$

Even though $\mathrm{Cov}(X, Y) = 0$, the variables are *not independent*, since Y is deterministically given by X:

$$Y = X^2.$$

For example,

$$P(Y = 0 \mid X = 0) = 1 \neq P(Y = 0) = \frac{1}{3}.$$

Hence, X and Y are dependent, even though their covariance is zero.

Covariance only captures **linear dependence**. Here the relationship between X and Y is nonlinear (parabolic), so their covariance cancels out to zero even though they are functionally related.

1.4.3 Correlation Coefficient of Two Random Variables

Let X, Y be two random variables with $E[X^2]$, $E[Y^2] < +\infty$.

The correlation coefficient (also known as Pearson's correlation coefficient) is defined as:

$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y},$$

where σ_X and σ_Y are the standard deviations of the variables X and Y.

Correlation gives us information about the dependency of two random variables removing dimensionality.

1.4.4 Properties of Covariance and Correlation

Let X, Y be two random variables with $E[X^2], E[Y^2] < +\infty$.

1. $\mathrm{Var}(X + Y) = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}(X, Y)$
2. $\rho(X, Y) \in [-1, 1]$

Proof

(i)

$$\begin{aligned}\mathrm{Var}[X + Y] &= E[(X + Y)^2] - E[X + Y]^2 \\ &= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2(E[XY] - E[X]E[Y]) \\ &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}(X, Y).\end{aligned}$$

(ii)

$$\begin{aligned}
0 &\leq \mathrm{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) \\
&= \mathrm{Var}\left(\frac{X}{\sigma_X}\right) + \mathrm{Var}\left(\frac{Y}{\sigma_Y}\right) \pm 2\,\mathrm{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\
&= \frac{\sigma_X^2}{\sigma_X^2} + \frac{\sigma_Y^2}{\sigma_Y^2} \pm 2\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} \\
&= 2 \pm 2\rho(X,Y).
\end{aligned}$$

Since the variance is always positive:

$$-1 \leq \rho(X,Y) \leq 1. \quad \square$$

1.4.5 Exercises

Calculate the correlation coefficient for the first problem presented(the dices problem)

Solution: We want to find $\rho(X_1, X_2)$ where $X_1 = D_1 + D_2$ and $X_2 = D_1$.

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}}$$

We need the covariance and the two standard deviations.

- **Covariance:**

$$\begin{aligned}
\mathrm{Cov}(X_1, X_2) &= \mathrm{Cov}(D_1 + D_2, D_1) \\
&= \mathrm{Cov}(D_1, D_1) + \mathrm{Cov}(D_2, D_1) \\
&= \mathrm{Var}(D_1) + 0 \quad \text{(since } D_1, D_2 \text{ are independent)}
\end{aligned}$$

  For a single die D_1:
  $E[D_1] = 3.5$
  $E[D_1^2] = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}$
  $\mathrm{Var}(D_1) = E[D_1^2] - (E[D_1])^2 = \frac{91}{6} - (3.5)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12}$
  So, $\mathrm{Cov}(X_1, X_2) = \frac{35}{12}$.

- **Standard Deviations:** $\sigma_{X_2} = \sigma_{D_1} = \sqrt{\mathrm{Var}(D_1)} = \sqrt{\frac{35}{12}}$
  $\mathrm{Var}(X_1) = \mathrm{Var}(D_1 + D_2) = \mathrm{Var}(D_1) + \mathrm{Var}(D_2) \quad \text{(independent)}$
  $\mathrm{Var}(X_1) = \frac{35}{12} + \frac{35}{12} = \frac{70}{12} = \frac{35}{6}\ \sigma_{X_1} = \sqrt{\frac{35}{6}}$

- **Correlation Coefficient:**

$$\rho(X_1, X_2) = \frac{35/12}{\sqrt{35/6} \cdot \sqrt{35/12}} = \frac{35/12}{\sqrt{\frac{35\cdot 35}{6\cdot 12}}} = \frac{35/12}{\frac{35}{\sqrt{72}}}$$

$$\rho(X_1, X_2) = \frac{35}{12} \cdot \frac{\sqrt{72}}{35} = \frac{\sqrt{72}}{12} = \frac{\sqrt{36 \cdot 2}}{12} = \frac{6\sqrt{2}}{12} = \frac{\sqrt{2}}{2}$$